Glass and Ceramics, Vol. 71, Nos. 3 – 4, July, 2014 (Russian Original, Nos. 3 – 4, March – April, 2014)

UDC 666.1:62-51

STATISTICAL DESCRIPTION OF TECHNOLOGICAL PROCESSES IN GLASS PRODUCTION

R. I. Makarov[1]

Translated from *Steklo i Keramika*, No. 4, pp. 14 – 16, April, 2014.

The particularities of multiple regression as a method for mathematically describing the technological processes in glass production are examined. The procedure for choosing the structure of the regression model based on analysis of the variation of the dependent variable under the conditions of normal flow of the process being analyzed is described. The efficacy of the procedure proposed for describing the technological process of glass ribbon formation in a float bath is shown.

Key words: sheet glass, technological process, statistical description.

Effective control of the technological processes in glass production is possible when the basic mechanisms of the process are represented mathematically. The technological processes in glass production are complex objects of control. The regime variables and quality parameters of the glass depend on numerous controllable and uncontrollable factors. For this reason obtaining a serviceable mathematical description based on statistical data on the functioning of an object under normal operating conditions is a complex research problem.

The literature on the practical application of multiple regression shows that in most cases researchers use a regression model with independent normal errors and equally accurate measurements to construct a statistical mathematical description of continually operating objects [1].

The construction of a statistical mathematical description of an object consists in finding relations between each of the output variables of the object and all other controllable input variables. The desired relation is sought on the basis of statistical data representing the results of measurements of the variables of the object in a normal operating regime.

To construct the exact regression equations it is necessary to know the conditional distribution law for the output parameters. In the practice of statistics such information usually cannot be obtained; instead, suitable approximations are sought for the function $f(x_1, x_2, ..., x_k)$ describing the dependence of a conditional average value of the output variable y on the prescribed values of the arguments $x_1, x_2, ..., x_k$. If the multidimensional random quantity $(y, x_1, x_2, ..., x_k)$ satisfies a $(k + 1)$-dimensional normal distribution law, then the regression equation of the output variable y with respect to the controllable input variables $x_1, x_2, ..., x_k$ is linear in x [2]. However, in the practice of statistics attention is confined to searching for suitable approximations for the unknown true regression function $f(x_1, x_2, ..., x_k)$, since the researcher does not know the exact probability distribution of the output variable y being analyzed for known values of the arguments $x_1, x_2, ..., x_k$.

If the wrong class is chosen for the regression function, the statistical results and estimates will not be consistent and increasing the number of observations will not give an estimate of y closer to the true regression function $f(x_1, x_2, ..., x_k)$. If the correct class is chosen for the regression function, the inaccuracy in the description of the true regression functions with the aid of the y is explained by the limited nature of the sample of the statistical observations [2]. The best reconstruction of the output variable of the object $y(x_1, x_2, ..., x_k)$ and the unknown regression function is obtained on the basis of the statistical data by the least-squares method.

In the regression method the choice of structure is one of the important problems having a significant effect on the accuracy of the model. There are only a few methods for choosing the structure of a regression model. In practice linear models are insufficient for describing real objects. For this reason, when choosing a structure for the model the degree of nonlinearity of the observational data is determined using the apparatus of the correlation theory of random functions. For a multidimensional object the degree of nonlinearity of the dependent variable y relative to $x_1, x_2, ..., x_k$ is determined as the difference of the squares of multivariate, mutually normalized, dispersion and correlation functions [3].

[1] A. G. and N. G. Stoletov Vladimir State University, Vladimir, Russia (e-mail: makarov.ruslan@gmail.com).

In practice the form of the regression equation is chosen on the basis of an analysis of the physical essence of the object being studied and the observational results.

We shall examine a procedure for choosing the structure of the regression model on the basis of an analysis of the dispersion (variation) of the output variable of the object being analyzed. The function $y(x_1, x_2, ..., x_k)$ can be represented graphically by the response surface in the $(k + 1)$-dimensional space of the arguments $x_1, x_2, ..., x_k$. Choosing a structure for the model means searching for a suitable approximation describing the dependence of the conditional average value of the output variable y on the prescribed values of the arguments $x_1, x_2, ..., x_k$. If the variation of the dependent variable y is very small in the observational sample, the conditional average value of the output variable y can be approximated with acceptable accuracy by a linear surface.

Such an approximation will bring the sample estimate y close to the true regression function $f(x_1, x_2, ..., x_k)$ if the variation of the dependence variable is significant. This makes it necessary to use polynomial regression to describe the data being analyzed. In most cases polynomials of degree no higher than second are used to describe stationary technological processes.

In constructing a statistical mathematical description the researcher is faced with the problem of choosing a linear or nonlinear multidimensional function. It is necessary to determine which function is the true one and to calculate the parameters of the model chosen.

In practice it is important to know how well the model corresponds to the object being described and whether or not the prediction accuracy should be increased. The value of the multiple correlation coefficient of the regression equation such that $R \geq 0.86$ could provide answers to these questions. The error of prediction based on the regression equation will be a factor of 2 smaller than the error of prediction based on the average value of the dependent variable y. Any measure that gives a very small increase in the multiple correlation coefficient beyond the limit $R = 0.86$ is justified in practice because the serviceability of the regression equation can increase considerably [1].

The following algorithm is proposed for choosing the structure of the regression equation.

1. Determine the coefficient of variation of the dependent variable y in the statistical sample used for constructing the model.

2. For a large coefficient of variation (> 20 – 30%) construct linear and nonlinear regression models. Check the adequacy of the models developed.

3. Compare the accuracy of the models developed according to the magnitude of the residual dispersion according to the Fisher criterion:

$$F = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)},$$

where s_1^2 and s_2^2 are the residual dispersions of the models being compared.

Compare the computed value of the criterion with the tabulated value for the significance level $\alpha = 0.05$ and number of degrees of freedom $(n_1 - k_1 - 1)$, $(n_2 - k_2 - 1)$. If the tabulated value of the criterion is greater than the computed value, then a model with the smaller residual dispersion is chosen. Otherwise the models are taken to be of equal accuracy and preference is given to the simpler model, viz., the linear model.

4. For multiple correlation coefficients $R < 0.86$ for the chosen model take measures to increase the accuracy of the model. Such measures could be increasing the size of the initial sample used to construct the regression model, including additional factors in the model, choosing a different structure for the model and others.

We shall now examine a mathematical description of the technological process leading to the formation of the glass ribbon in a float tank. Optical distortions of the glass are largely determined by the formation regime [4]. The specifications given in GOST 111–2001 (Sheet Glass: Technical Conditions) do not allow distortions of 'zebra' bands for M0 grade glass at angles ≤ 50°. For optical distortions visible in reflected light for M0 grade glass, distortions of the index of a reflected raster greater than 3 mm are not allowed.

A sample comprised of 365 measurements of the average daily indices of optical distortions of the sheet glass being produced and the technological operating regimes of the float bath was used to construct a regression model. The following were studied as the influential variables: the temperature in the passages in the float bath, daily fluctuations in the glass density and the change in the thickness of the glass produced. The quality indices for glass formation were taken to be the optical distortions visible in transmitted light ('zebra') and distortions visible in reflected light (raster).

The coefficients of variation of the optical distortions of the glass in the statistical sample were 11.8% in terms of the zebra index and 63.5% in terms of the raster. The coefficient of variation of the reflected raster index is significant, and for this reason an analysis using a nonlinear regression model is desirable.

Linear and nonlinear models were constructed and assessments of the models are presented in Table 1. As one can see from the tabulated data a quadratic polynomial regression model gives a more accurate description of the dependence of the deviations of the index of the reflected raster on the formation regime and the thickness of the glass ribbon formed. The effect of all other factors was found to be statistically insignificant. The nonlinear regression equation describing the deviations of the reflected raster index Ra has the form

$$Ra = -455.57 - 4.91\delta + 0.938\theta_1 + 0.57\delta^2 - 0.00047\theta_1^2,$$

where δ is the thickness (in mm) of the glass produced and θ_1 is the temperature in the first passage of the float bath, °C.

TABLE 1. Estimates of Raster Regression Models

Index	Structure of the model	
	linear	quadratic polynomial
Number of factor variables	4	2
Multiple coefficient of regression	0.53	0.77
Significance of the multiple coefficient of regression	Significant	Significant
Residual dispersion of the model, mm^2	8.7	0.64

The coefficient of variation of the optical distortions of the glass is negligible (11.8%) in terms of the zebra index in the statistical sample. This made it possible to give an adequate description of the process of linear regression in terms of a model of the form

$$\text{'zebra'} = -34.97 + 5.91\delta + 0.118\theta_1 - 0.085\theta_{20},$$

where θ_1 and θ_{20} are the temperatures in the first and 20th passages of the float bath, °C.

The multiple correlation coefficient of the model equals $R = 0.84$ and the residual dispersion is 11 $(°C)^2$. The use of a quadratic polynomial structure made it impossible to increase the accuracy of the model of optical distortions of the glass in terms of the 'zebra' index.

If necessary, the accuracy of the models developed can be increased by including in the model additional factors affecting the process resulting in the formation of a glass ribbon in the float-bath.

Analysis of the variation of the indices of the technological process resulting in the formation of a glass ribbon in a float bath made it possible to choose regression models that adequately describe the dependence of the optical distortions on the glass thickness and the formation regime.

REFERENCES

1. V. P. Borodyuk and É. K. Letskii, *Statistical Description of Industrial Objects* [in Russian], Énergiya, Moscow (1971).
2. A. M. Dubrov, V. S. Mkhitaryan and L. I. Troshin, *Multidimensional Statistical Methods* [in Russian], Finansy i Statistika, Moscow (2006).
3. F. F. Pashchenko, *Introduction to Valid Methods for Modeling Systems, Pt. 1, Mathematical Principles of Systems Modeling* [in Russian], Finansy i Statistika, Moscow (2006).
4. R. I. Makarov, E. R. Khorosheva and S. A. Lukashin, *Automation of Sheet Glass Production (Float Method)* [in Russian], Assotsiatsiya Stroitel'nykh Vuzov, Moscow (2002).